DOCUMENT RESUME

ED 182 762 CS 205 823

AUTHOR Keech, Catharine
TITLE How to Use the Method of Adjusted Averages.
PUB DATE [79]
NOTE 17p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment: *Statistical Analysis: *Student Evaluation: *Writing Skills
IDENTIFIERS *Bay Area Writing Project

ABSTRACT

The method of adjusted averages is presented in this paper as an easily applied statistical procedure useful in comparing average scores of two or three groups after an instructional period. Step-by-step preparations for using the method are described, including gathering scores from pre- and postinstruction measures of ability, dividing scores into strata, charting the strata, figuring averages, and using scores from an imaginary population to predict mean scores. Instructions for calculating the adjusted averages of the ninth grade classes of two sample schools are presented in detail, along with information on how to interpret the results. Other applications of the procedure that would be useful to teachers and evaluators are suggested. (AEA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Bay Area Writing Project
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED182762

HOW TO USE THE METHOD OF ADJUSTED AVERAGES

by Catharine Keech

The method of adjusted averages is a relatively straightforward, easy to apply statistical procedure. It is useful when we wish to compare the average scores of two or three groups to see which instructional program seems to yield the best performance at the end of some instructional period.

Suppose we gave the same writing test to students at three schools within one district in the spring of one year. We find that School Three students have an average score of 56.7, while School Two students average only 52.4, and School One looks really bad, with an average score of 45.3. On the basis of these scores, the district superintendent, backed by the school board, awards a certificate of merit with an extra vacation bonus to the teachers at School Three. School Two teachers are told to "keep up the good work." School One teachers are told to "shape up or ship out," and a massive curriculum revision is planned which requires teachers to attend summer workshops without pay.

Under some circumstances, the school administration might be justified in its conclusions, if not in the actions taken. On the face of the evidence, School Three teachers appear to be doing a much better job than those in School Two, and School One teachers appear to be having serious trouble. But a teacher from School One speaks up and says, "Wait a minute. I think we have the best staff in the district. But we also have the poorest students--in every sense. They have the poorest school attendence record before they come to us, their parents have the lowest level of education in the district, and their verbal skills when they enter are well below the rest of the district. Sure, we have a few top ability kids, but not as many as other schools."

The School One teacher happens to teach ninth grade English. He asks a district statistician to compare the performance of his ninth graders on the writing test to the performance of ninth graders from the other schools. He asks the statistician to use some method which would take into account the

CS205423

initial differences in ability and achievement between his students and the others.

The district statistician considers a number of approaches to the problem. He knows that all the ninth graders had been tested in the fall using a standardized verbal skills test. Although this test was not an adequate test of writing skills, it did measure some of the abilities that seemed to affect general performance in school, including performance on writing assignments. He has several alternatives:

1. He can perform an analysis of covariance using the fall test scores as the covariate or input variable. But to use this method properly, he has to make certain assumptions about the data that he is not sure he can justify. He must assume normal distribution of scores on both measures, a common slope for the regression lines, and common variation about the regression lines. Given the data he has, these assumptions are suspect.

2. He can perform a two-way analysis of variance on the data, with the two factors affecting output being school and level of initial ability of students. To do this, he would establish four or five levels of ability based on the fall test. He would then test for differences between school means at each individual level of ability. In this way, it wouldn't matter if one school had many more lower level students than another; as long as the lower level students at both schools performed about the same on the writing test, he would be able to show that the school program was not responsible for the differences between the three schools in over-all mean scores.

   The major weakness of this method is that, though statistically sound, it generally leads to unusable or uninformative results. This is especially the case when, in each school, there is a small number of students at each ability level.

3. The statistician decides to use a method which was originally developed by researchers in the biomedical sciences and is now universally accepted in public health, government, and medicine. It is useful in education for comparing groups which are not randomly constructed by the researcher, but must be taken as they exist. The model used is called the METHOD OF ADJUSTED AVERAGES.* Because it is simple and intuitive, the statistician taught the teacher the procedure as he analyzed the data.

---

*METHOD OF ADJUSTED AVERAGES is described in Marascuilo, L., and Dagenais, F., "Measuring Differences Among Non-Randomized Groups--an Epidemiological Model for Identifying Successful School Programs," Journal of Experimental Education, Fall, 1979.

The statistician said, "Let's take one ninth grade class from each school, and I will show you how to apply the method of adjusted averages. We will calculate WHAT EACH CLASS AVERAGE WOULD BE IF ALL THREE CLASSES HAD THE SAME INITIAL MAKE-UP OF STUDENTS at different ability levels as determined by their level of achievement on the fall standardized test."

PREPARATION

STEP ONE: Find the numbers of students and the actual writing test averages of each class of ninth graders.

	n (number)	$\bar{X}$ (mean score)
School One:	30	45.3
School Two:	39	52.4
School Three:	30	56.7

STEP TWO: Find each student's individual score on the fall basic skills test, verbal section (or any pre-instruction measure of ability you have available). This measure should correlate relatively well with the writing test measure ($r = .50$ or better).

Divide the possible scores on this test into four, five, or more strata. If the test is standardized, you may use the test maker's divisions into quartiles.

Bottom quartile: Students scoring 39.5 or less

Second quartile: Students scoring between 39.5 and 49.5

Third quartile: Students scoring between 49.5 and 59.5

Top quartile: Students scoring 59.5 or more

(Since the scores are reported as whole numbers (39, 40...49, etc.) there is no danger that any student will score 49.5 or 59.5. These imaginary scores with half points prevent confusion about which real scores fall into which brackets.)

STEP THREE: Make a chart like Chart I following. In the first column list the four (or more) strata determined by scores on the fall test. Then for each school, list the number of students in each strata.

For instance, in the bottom strata (scores of 39.5 or less), School One has twelve students, School Two has ten, and School Three has only two students. In the next level up, (scores of 39.5 to 49.5) School One has eight students, School Two, three students, and School Three has six students, and so on.

STEP FOUR: Find the average score on the SPRING writing test for the students in each ninth grade class who are in the bottom strata, based on their fall testing.

For School One, the twelve students in the bottom strata have an average score of 35 on the Spring writing test.

I

METHOD OF ADJUSTED AVERAGES

Three Schools with Different Raw Averages But Identical Adjusted Averages

Input based on Fall Testing	School One Spring Scores		School Two Spring Scores		School Three Spring Scores		Standard Population for $\bar{N}$ (50, 100)	
	n_h	$\bar{X}_h$	n_h	$\bar{X}_h$	n_h	$\bar{X}_h$	n_h	$\bar{X}_h$
$X \leq 39.5$	12	35	10	35	2	35	16	35
$39.5 \leq X \leq 49.5$	8	45	3	45	6	45	34	45
$49.5 \leq X \leq 59.5$	7	55	8	55	7	55	34	55
$59.5 \leq X$	3	65	14	65	15	65	16	65
Total/Average	30	45.3	39	52.4	30	56.7	100	50

STEP FIVE: Repeat the procedure of Step Four for each of the strata. Calculate the mean scores for the ninth grade students in each strata at each school.

STEP SIX: Create an imaginary population in the last column of your chart. The imaginary population may be made up of a "standard group," with the following distribution:

lowest strata	16	
second strata	34	Total = 100
third strata	34	
highest strata	16	

Alternatively, the imaginary population may reflect the proportion of students performing in each strata on the Fall test.

lowest strata	25	
second strata	25	Total = 100
third strata	25	
highest strata	25	

Since the numbers in each strata in the imaginary population are wholly arbitrary, you may even assign numbers that will make your final computations easier, such as:

lowest strata	10	
second strata	10	
third strata	10	Total = 40
highest strata	10	

STEP SEVEN: Find the mean score on the Spring Writing Test of ALL the students in the lowest strata, regardless of school. Assign this mean score to the lowest strata students in the IMAGINARY POPULATION.

In this case, since all the schools had the same mean score for lowest strata students, 35 will also be the mean score of the lowest strata students in the imaginary population. We think of this score as the predicted mean score for students in this strata.

Calculate the predicted mean score for each strata by finding the mean score for all students in each strata, regardless of school. Level two students have a mean score of 45; level three students have a mean score of 55; and level four students have a mean score of 65. Enter these means in the chart.

STEP EIGHT: Find the mean score for the entire IMAGINARY POPULATION. Since the score at each level or strata is the predicted score for that strata, we may think of the mean score for this population as the predicted mean score FOR ANY POPULATION WHICH HAS THE SAME DISTRIBUTION OF STUDENTS as the standard or imaginary population. In other words, if our imaginary population has an equal number of students at each level, then any class or group which has an equal number of students at each level IS PREDICTED TO have a group mean which is like the group mean of our standard or imaginary population.

All of our subsequent calculations are based on this assumption. We are trying to find out what each schools' ninth grade scores would be IF each school had the same number of ninth graders as every other school at each level of ability.

In this case, we can see at a glance that the adjusted averages will be the same for all schools. If the students at a given level have the same mean score as other students at the same level regardless of which school they attend, we may conclude that if each school had the same number of students at each level, each school would have the same over-all mean performance on the spring writing test, which in this case, will be 50.

Unfortunately, life is never so neat as in this sample chart. To demonstrate how we calculate the adjusted averages for two more dissimilar groups, and how we test whether the resulting averages are statistically different from one another, let us consider a slightly different case, which appears in Chart Two.

METHOD OF ADJUSTED AVERAGES

Comparison of Two Schools

Input based on Fall Testing	School One Spring Scores			School Two Spring Scores			Standard Population
	n_h	$\bar{X}_h$	S_h	n_h	$\bar{X}_h$	S_h	$\bar{n}_h$
$X \leq 39.5$	19	38.6	2.1	6	37.5	1.3	16
$39.5 \leq X \leq 49.5$	22	47.1	3.1	12	46.2	2.2	34
$49.5 \leq X \leq 59.5$	13	56.3	1.7	23	59.2	3.6	34
$59.5 \leq X$	3	62.9	.8	10	68.4	4.9	16
Total	57	47.3		51	55.4		100

School One :

$$\bar{X}_{1}^{A} = \frac{16(38.6) + 34(47.1) + 34(56.3) + 16(62.9)}{100} = 51.4 \text{ (adjusted mean)}$$

$$SE^{2}_{\bar{X}_1 A} = (.16)^2 \frac{(2.1)^2}{19} + (.34)^2 \frac{(3.1)^2}{22} + (.34)^2 \frac{(1.7)^2}{13} + (.16)^2 \frac{(.8)^2}{3} = .0874$$

School Two :

$$\bar{X}_{2}^{A} = \frac{16(37.5) + 34(46.2) + 34(59.2) + 16(68.4)}{100} = 52.8 \text{ (adjusted mean)}$$

$$SE^{2}_{\bar{X}_2 A} = (.16)^2 \frac{(1.3)^2}{6} + (.34)^2 \frac{(2.2)^2}{12} + (.34)^2 \frac{(3.6)^2}{23} + (.16)^2 \frac{(4.9)^2}{10} = .1804$$

H_0: $\mu = 50$

$$Z_1 = \frac{51.4 - 50}{\sqrt{.0874}} = \frac{1.4}{.296} = 4.74^*$$

$$Z_2 = \frac{52.8 - 50}{\sqrt{.1804}} = \frac{2.8}{.425} = 6.54^*$$

H_0: $\mu_1 = \mu_2$

$$Z = \frac{51.4 - 52.8}{\sqrt{.0874 + .1804}} = \frac{-1.4}{\sqrt{.2678}}$$

$$= \frac{-1.4}{.517} = -2.71^*$$

*Significant at $p \leq .05$

CALCULATING THE ADJUSTED AVERAGES

In this case, let us compare the spring test scores of all ninth graders at a public school (School One) with the scores of all ninth graders at a private school (School Two). Again, we use fall standardized test scores and set up a chart, with an imaginary population, exactly as in STEPS ONE through EIGHT above. (See Chart Two)

STEP NINE: Find the adjusted average for School One and School Two.

a. Use the numbers of students at each level in the IMAGINARY or STANDARD POPULATION. (In this case: 16, 34, 34, 16.)

b. Multiply the standard population number at each level times the mean score of ninth graders in School One at each level and add the results:

$$16(38.6) + 34(47.1) + 34(56.3) + 16(62.9) = 5140$$

c. Divide the result by the total number of students in one imaginary, or standard, population:

$$\bar{X}_{A1} = \frac{5140}{100} = 51.4$$

The adjusted average score for ninth grade students in School One is 51.4.

d. Repeat the process for School Two:

$$\bar{X}_{A2} = \frac{16(37.5) + 34(46.2) + 34(59.2) + 16(68.4)}{100} = 52.8$$

The adjusted average score for ninth grade students in School Two is 52.8.

These two adjusted averages are much closer than the original raw averages. But they are not identical. We need to discover whether the difference between them is only what might be expected, given the general variation in scores on this test and the numbers of students in the sample. To do this, we must test for the statistical significance of the differences between the scores.

STATISTICAL COMPARISONS BETWEEN SCHOOLS OR PROGRAMS

To identify successful or unsuccessful teaching programs in the two schools, we may use standard statistical procedures for comparing the adjusted averages of the two schools. The null hypothesis that there is no statistically significant difference between the adjusted means of the two schools can be stated:

$$H_0: \ mu_1 = mu_2$$

(Since the sample sizes are relatively large, it is known that the sampling distribution of $\bar{X}^A$ tends to a normal form as a result of the Central Limit Theorem.)

We may test this hypothesis by finding the z value of the contrast between schools:

$$Z = \frac{\bar{X}^A_1 - \bar{X}^A_2}{\sqrt{SE^2_{\bar{X}^A_1} + SE^2_{\bar{X}^A_2}}}$$

The null hypothesis is rejected if $Z > Z_{\alpha/2=.025}$ or $Z < -Z_{\alpha/2=.025}$. In this case, $Z_{\alpha/2=.025}$ at $p \leq .05$ is ± 1.96

We have found the adjusted averages of the two schools:

$$\bar{X}^A_1 = 51.4 \quad \text{and} \quad \bar{X}^A_2 = 52.8.$$

Now we must calculate the STANDARD ERROR (squared) of the adjusted mean $(SE^2_{\bar{X}^A})$ for each school:

For each school: $$SE^2_{\bar{X}^A} = \sum_{k-1}^{K} \left[\frac{N_k}{N}\right]^2 \frac{s^2_k}{n_k}$$

Where: N = the total number in the standard or imaginary population

N_k = the number in the standard population in strata k (each individual strata, calculated separately)

s^2_k = sample variance in strata k, for the individual school

n_k = the number in strata k, for the individual school

For School One, the standard error squared is found (see Chart II):

$$SE^2_{\bar{X}^A_1} = (.16)^2 \frac{(2.1)^2}{19} + (.34)^2 \frac{(3.1)^2}{22} + (.34)^2 \frac{(1.7)^2}{13} + (.16)^2 \frac{(.8)^2}{3} = .0874$$

For School Two, the standard error squared is found:

$$SE^2_{\bar{X}^A_1} = (.16)^2 \frac{(1.3)^2}{6} + (.34)^2 \frac{(2.2)^2}{12} + (.34)^2 \frac{(3.6)^2}{23} + (.16)^2 \frac{(4.9)^2}{10} = .1804$$

Calculating the value for z:

$$z = \frac{51.4 - 52.8}{.0874 + .1804} = \frac{-1.4}{.2678} = -2.71^*$$

*significant at $p \leq .05$.

Since -2.71 is further from the value of zero (or no difference) than the test statistic, $z = \pm 1.96$, we reject the null hypothesis of no difference between the schools.

INTERPRETING RESULTS

In this case, we must conclude that some factors in the school setting other than, or interacting with, the initial ability of students as measured on the fall verbal test are affecting their performance on the writing test. The difference in performance on the writing test has been shown to be statistically significant: that is, unlikely to have been caused by random factors in the measurement. This does not mean that the difference is educationally significant. The teachers in the two schools, together with the administration, would have to decide, upon studying the writing sample, whether a difference of 1.4 points is likely to matter to students in the real world. In this case, a difference of 1.4 points is unlikely to move students out of one ability strata performance into another. On the other hand, the difference in a writing program that produced students who scored a minimum of one third of a standard deviation higher than another group, even when average scores were adjusted, may be an important difference in the long run.

Finally, teachers in the two schools must carefully examine reasons for the difference in performance between the schools. An enriched, or more appropriately directed program in School One could possibly raise the scores of its upper level students. From the chart, we can see that, while School One's lower level students perform slightly better than those in School Two, the upper level students in School One seem to be suffering. Since there are fewer of these students, perhaps School One teachers are concentrating their attention on the more apparently needy lower level students who make up the majority of the class. On the other hand, perhaps the level of discussion in the School One class is slightly less stimulating because of the different distribution of students. Perhaps the higher level students are not spurred to competitive endeavor.

It should be possible, with the information provided by the analysis, to reexamine the school program for these students with a view to correcting deficiencies. Administrators should, of course, not use this comparison of the two ninth-grade programs to generalize to other school programs. The method of adjusted averages must be applied to the entire school curriculum to determine how the over-all program compares to School Two's program.

OTHER APPLICATIONS OF THE PROCEDURE

When comparing two or more schools, the analysis can be done separately for each grade level. Or, schools can use a fall test to stratify the populations regardless of the grade level of students. Finally, schools can be compared using grade levels themselves as the input variable for stratifying the population.

In addition to comparing schools to one another, each school can be separately compared to the predicted mean, or the average score that would be produced by a standard, or imaginary population.

To calculate the expected mean for the entire population (when this is not provided by the test maker, as in standardized tests), first find the mean

at each strata for all students regardless of schools:

$$\bar{X}_{s1} = \frac{19\ (38.6) + 6\ (37.5)}{19 + 6} = 38.34$$

$$\bar{X}_{s2} = \frac{22\ (47.1) + 12\ (46.2)}{22 + 12} = 46.78$$

$$\bar{X}_{s3} = \frac{13\ (56.3) + 23\ (59.2)}{13 + 23} = 24.11$$

$$\bar{X}_{s4} = \frac{3\ (62.9) + 10\ (68.4)}{3 + 10} = 67.13$$

Then weight the mean at each strata using the number of students at each strata in the standard population, and divide by the total number in the standard population for the expected mean:

$$(E)\ \bar{X} = \frac{16\ (38.34) + 34\ (46.78) + 34\ (24.11) + 16\ (67.13)}{100} = 40.98$$

In this case, the best estimate of mean score we might expect from all schools, if they had similar distributions of population, would be

$$(E)\ \bar{X} = 40.98$$

Although the highest and lowest school may have a statistically significant difference in mean scores, it may also be true that neither school's adjusted mean is statistically significantly different from the expected mean for the whole group.

Teachers and evaluators should be aware that reports of "no statistically significant difference" in adjusted averages may mean quite different things, depending on which hypothesis is being tested:

H_0: Adjusted Average $School_1$ - Adjusted Average $School_2$ = 0

as opposed to:

H_0: Adjusted Average $School_1$ - Expected Average = 0

Adjusted Average $School_2$ - Expected Average = 0